Materialize

Materialize + dbt: Streaming for the modern data stack

Jessica Laughlin
Materialize

It’s the modern data stack’s world, we just live in it.

First, some definitions

data stack → collection of all your data tools

First, some definitions

data stack	→	collection of all your data tools
modern data stack	→	collection of all your data tools, in the cloud

What’s so good about the modern data stack?

- More powerful
- Less expensive

What’s so good about the modern data stack?

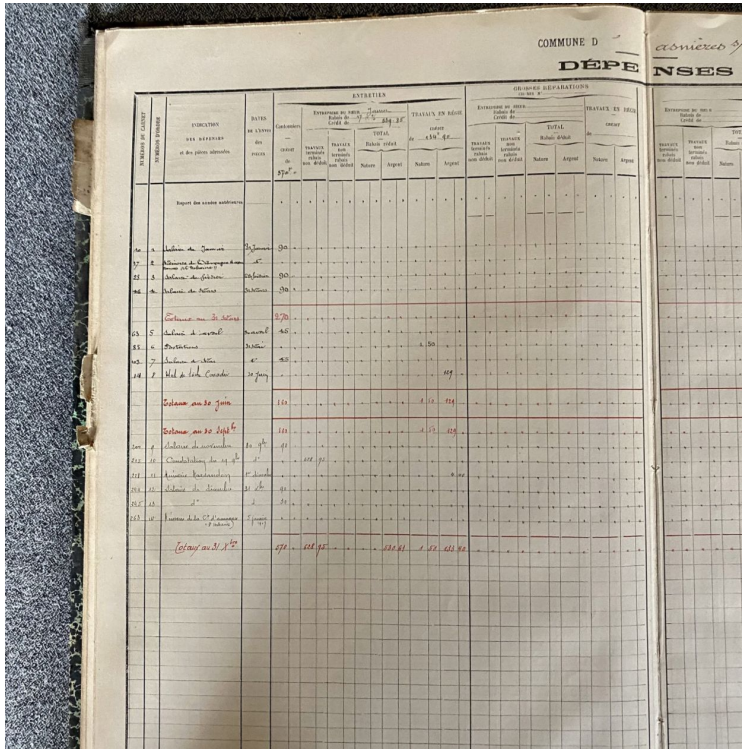
- More powerful
- Less expensive
- **SQL interface**

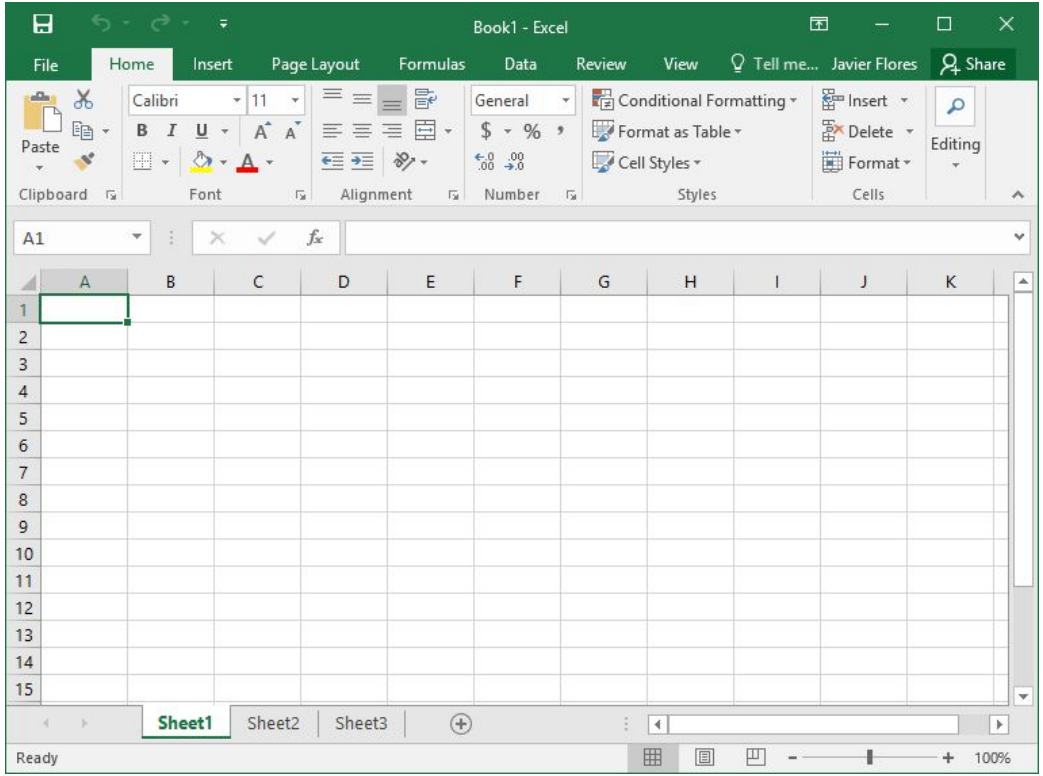
A small detour: interfaces are important.

COMMUNE D' asnières s/

DÉPENSES

Numéros du carnet	Numéros d'ordre	Indication des dépenses et des pièces adressées	Dates de l'envoi des pièces	ENTRETIEN — Cantonniers — Crédit de 570f	ENTRETIEN — Entreprise du sieur Jouanne? Rabais de 17 % Crédit de 539.35 — Travaux terminés rabais non déduit	Travaux non terminés rabais non déduit	Total — Rabais réduit — Nature	Total — Rabais réduit — Argent	ENTRETIEN — Travaux en régie — Crédit de 134f 90 — Nature	Travaux en régie — Argent	GROSSES RÉPARATIONS Chemin N° — Entreprise du sieur … Rabais de … Crédit de … — Travaux terminés rabais non déduit	Travaux non terminés rabais non déduit	Total — Rabais déduit — Nature	Total — Rabais déduit — Argent	Travaux en régie — Crédit de … — Nature	Travaux en régie — Argent	Entreprise du sieur … Rabais de … Crédit de … — Travaux terminés rabais non déduit	Travaux non terminés rabais non déduit	Total — Rabais … — Nature
		Report des années antérieures																	
10	1	Salaire de Janvier	31 Janvier	90															
17	2	Mémoire de [illegible] (L'Urbaine?)	d°																
23	3	Salaire de février	28 février	90															
46	4	Salaire de Mars	31 Mars	90															
		Totaux au 31 Mars		270															
63	5	Salaire d'avril	30 avril	45															
85	6	Prestations	31 Mai						1 50										
103	7	Salaire de Mai	d°	45															
14	8	Hel de Couradi?	30 Juin							129									
		Totaux au 30 Juin		360					1 50	129									
		Totaux au 30 Septbre		360					1 50	129									
200	9	Salaire de novembre	30 9bre	90															
215	10	Constatation du 19 9bre	d°		628 95														
228	11	[illegible]	1er décembre							4 40									
244	12	Salaire de décembre	31 Xbre	90															
245	13	d°	d	30															
263	14	Réunion de la Cie d'assurance (L'Urbaine)	5 janvier 1907																
		Totaux au 31 Xbre		570	628 95			530 61	1 50	133 40									


COMMUNE D
DÉPE
NSES


Book1 - Excel
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Tell me...
Javier Flores
Share
Paste
Clipboard
Calibri
11
Font
Alignment
General
Number
Conditional Formatting
Format as Table
Cell Styles
Styles
Insert
Delete
Format
Cells
Editing
A1
Sheet1
Sheet2
Sheet3
Ready
100%


LA.COM/CYBERTRUCK


PEPE
SILVIA

What’s wrong with the modern data stack?

The modern data stack is batch-based because streaming tools haven’t exposed that same SQL interface.

“Batch is good enough.”

pedram.sql
@pedram_navid

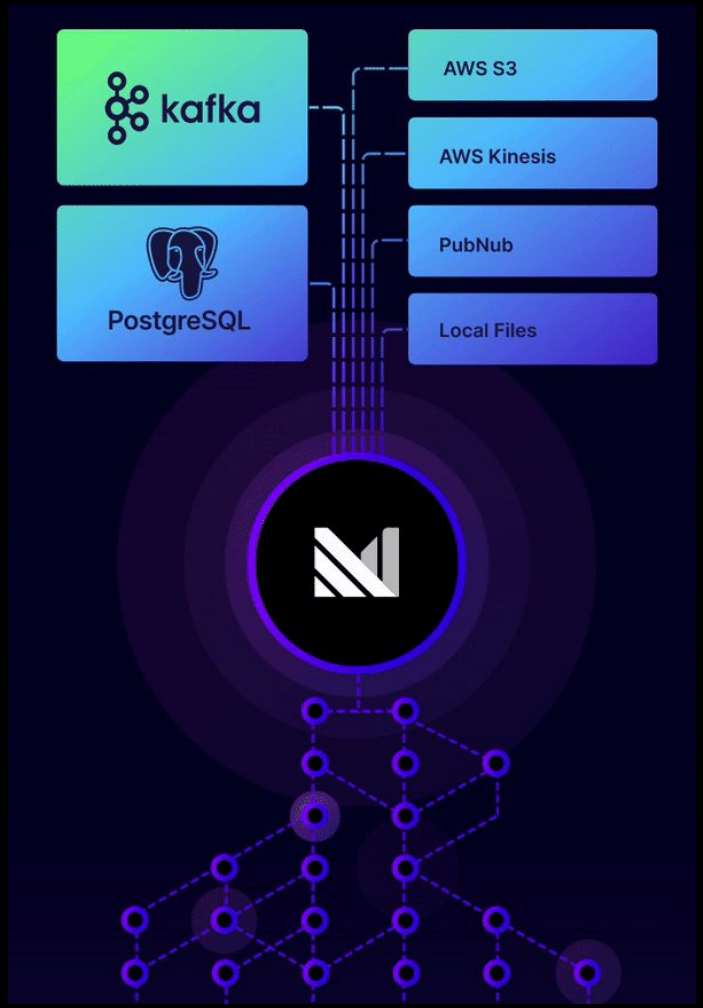
Replying to @sarahcat21

streams seem to me like one of those beautiful in principle and awful in practice things because of how hard it is to get right. i just can't imagine a 30-50 person company with one or two data hires running off streams but very easy to do off tables.

5:21 PM · Mar 15, 2022 · Twitter for iPhone

1 Retweet **24** Likes

Say hello to Materialize.


kafka
PostgreSQL
AWS S3
AWS Kinesis
PubNub
Local Files

Timely Dataflow and Differential Dataflow

Timely Dataflow ⟶ “a low-latency cyclic dataflow computational model, introduced in the paper Naiad: a timely dataflow system”

Timely Dataflow and Differential Dataflow

Timely Dataflow ⟶ “a low-latency cyclic dataflow computational model, introduced in the paper Naiad: a timely dataflow system”

Differential Dataflow ⟶ “a computational framework built on top of timely dataflow intended for efficiently performing computations on large amounts of data and *maintaining* the computations as the data change”


DON'T
PANIC!
1 2 3 4 5 6 7 8 9 0
Q W E R T Y U I O P
A S D F G H J K L
Z X C V B N M

You can build correct, real-time applications and analytics, just using SQL + Materialize.

Materialize’s superpower: incrementally-maintained materialized views.

Materialized views, a primer

What's a materialized view?	A database object that contains the result of a query.
Why would I use one?	Each time you query your database you will incur a cost. Materialized views cache your results, limiting your costs and keeping your results "fresh enough."
What's the downside?	Materialized views get stale, and refreshing them can be slow and expensive.

Incrementally-maintained materialized views

What's the same?

A database object that contains the result of a query.

What's different?

These materialized views update their results *incrementally* as the underlying data changes. This means they are never stale, and will always return correct, up-to-date results. Even better, this means they will only perform the necessary work.

COMMUNE D asnières s/

DÉPENSES

Numéros du carnet	Numéros d'ordre	Indication des dépenses et des pièces adressées	Dates de l'envoi des pièces	ENTRETIEN — Cantonniers — Crédit de 570f.	Entreprise du sieur Jeanneau, Rabais de 17 1/2 %, Crédit de 559.85 — Travaux terminés rabais non déduit	Travaux non terminés rabais non déduit	Total — Rabais réduit — Nature	Total — Argent	Travaux en régie — Crédit de 134f 90 — Nature	Travaux en régie — Argent	GROSSES RÉPARATIONS — Entreprise du sieur … Rabais de … Crédit de … — Travaux terminés rabais non déduit	Travaux non terminés rabais non déduit	Total — Rabais déduit — Nature	Total — Argent	Travaux en régie — Crédit de … — Nature	Travaux en régie — Argent
		Report des années antérieures														
10	1	Salaire de Janvier	31 Janvier	90												
17	2	Mémoire de la Compagnie d'assurances (l'Urbaine)	d°													
23	3	Salaire de février	28 février	90												
46	4	Salaire de Mars	31 Mars	90												
		Totaux au 31 Mars		270												
63	5	Salaire d'avril	30 avril	45												
85	6	Prestations	31 Mai						1 50							
103	7	Salaire de Mai	d°	45												
114	8	Etat de tâche Conradin	30 Juin							129						
		Totaux au 30 Juin		360					1 50	129						
		Totaux au 30 Septbre		360					1 50	129						
207	9	Salaire de novembre	30 9bre	90												
215	10	Constatation du 19 9bre	d°		628 95											
228	11	Quincaillerie Marsaudon	1er décembre							4 40						
244	12	Salaire de décembre	31 Xbre	90												
245	13	d°	d°	30												
263	14	Mémoire de la Cie d'assurances (l'Urbaine)	5 janvier 1907													
		Totaux au 31 Xbre		570	628 95			530 61	1 50	133 40						

Customer	**Item**	**Amount**
Jean	Some bread	4
Marie	Most bread	7
Pierre	Bread	5
		= 16

Customer	Item	Amount
Jean	Some bread	4
Marie	Most bread	7
Pierre	Bread	5
Louise	Tiny bread	3
		= 16

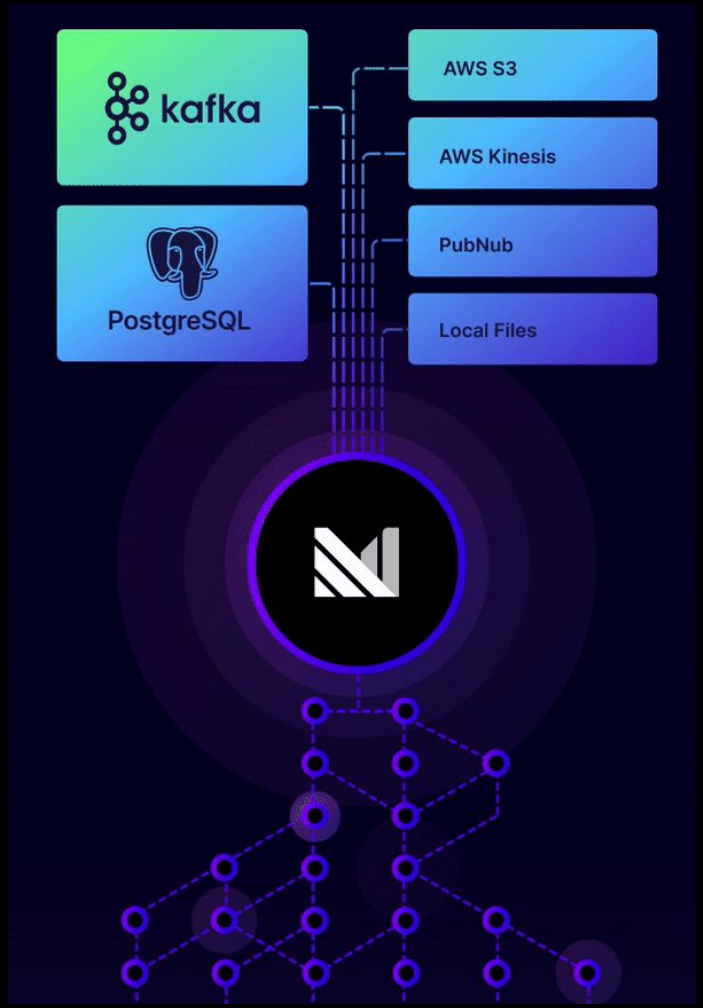
How would you calculate the updated sum?

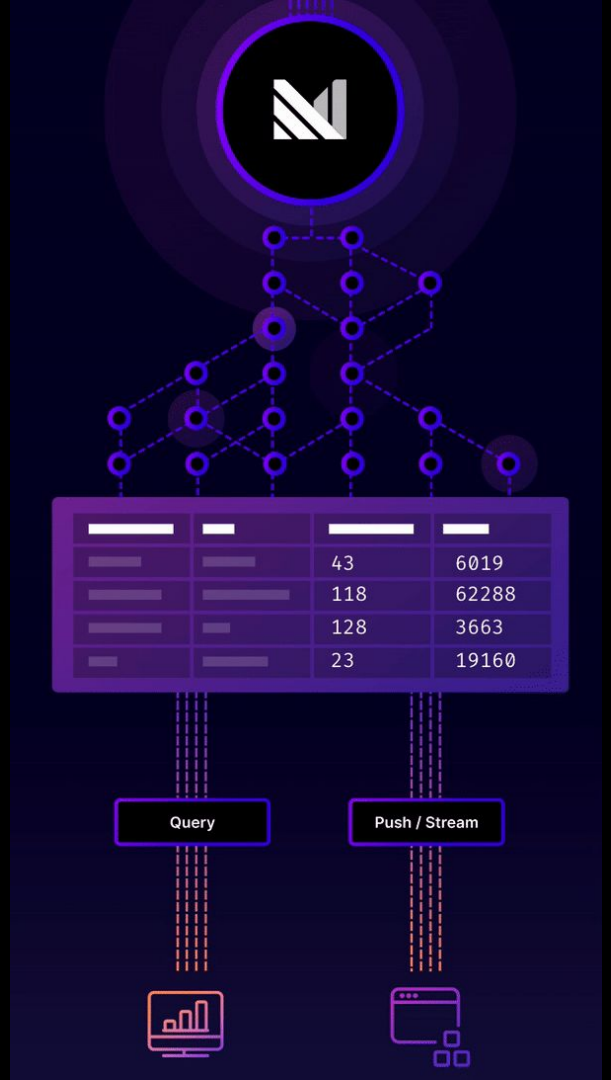
4 + 7 + 5 + 3 or 16 + 3?

Customer	Item	Amount
Jean	Some bread	4
Marie	Most bread	7
Pierre	Bread	5
Louise	Tiny bread	3
		= ~~16 + 3~~ = 19

How would you calculate the updated sum?

4 + 7 + 5 + 3 or **16 + 3**?


kafka
PostgreSQL
AWS S3
AWS Kinesis
PubNub
Local Files


43
6019
118
62288
128
3663
23
19160
Query
Push / Stream

Materialize + dbt

Because Materialize speaks SQL, you can transform your data in real-time using Materialize + dbt.

What's the same?

Model definitions	→	Because Materialize speaks PostgreSQL, you can define your models like you would with any other data warehouse.
dbt commands*	→	Use all of the same, familiar dbt commands over streams: dbt run, dbt test, dbt docs, and more.
Documentation, lineage	→	Document your streaming data and get handy lineage information, as usual.

What’s different?

Materialization types ⟶ You’ll want to use our custom “**materializedview**” materialization to create an incrementally-maintained materialized view.

What’s different

Materialization types	→	You’ll want to use our custom “**materializedview**” materialization to create an incrementally-maintained materialized view.
How some dbt commands are used	→	Let’s dig in, you’re going to be pleasantly surprised!

“dbt run” your models once and never again.

dbt test works on streaming data, too!

Run dbt test on your materialized views to catch data quality issues in real time.

dbt test + alert on streaming data #6

jwills started this conversation in **Show and tell**

jwills 14 days ago

edited

The repo for my hackday exercise is here: https://github.com/jwills/mz-hack-day-2022

Mostly copy-pasting from the Slack message where I described what I wanted to do:

"So one of the first things we do in our daily dbt run for our DWH at WeaveGrid is the standard staging work of renaming columns, cleaning up types, etc. followed by a bunch of sanity check dbt tests on that lightly processed staging data to ensure that there aren't any red flags in there that would mess up downstream table materializations. In my dream world, I would like to move that staging/testing work upstream- out of Snowflake and into dbt+materialize- so that I could run those tests "continuously" (read: with a cron job that executed like every 15 minutes-to-an-hour) to catch upstream data quality issues earlier, during business hours, and not after midnight UTC when a *&!^?# data quality issue is going to ruin my evening.

pedram.sql
@pedram_navid

Replying to @sarahcat21

streams seem to me like one of those beautiful in principle and awful in practice things because of how hard it is to get right. i just can't imagine a 30-50 person company with one or two data hires running off streams but very easy to do off tables.

5:21 PM · Mar 15, 2022 · Twitter for iPhone

1 Retweet **24** Likes

Jessica Laughlin @JLDLaughlin · Mar 15

you would be able to if the streaming system was hosted and providing the right interface - SQL! totally doable with a single dev.

2 6

pedram.sql
@pedram_navid

Replying to @JLDLaughlin and @sarahcat21

one day i will try this and if you live up to your promise i don’t know what i will do but it’ll be something

7:56 PM · Mar 15, 2022 · Twitter for iPhone

Thank you!

Questions? Come to our booth or find me at @JLDLaughlin or jessica@materialize.com

We're hiring! jobs@materialize.com